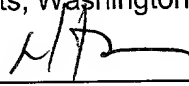


EXPRESS MAIL NO.: EL 909639427 US	DATE OF DEPOSIT: 2/22/02
This paper and fee are being deposited with the U.S. Postal Service Express Mail Post Office to Addressee service under 37 CFR §1.10 on the date indicated above and is addressed to: Box PATENT APPLICATION, Commissioner for Patents, Washington, D.C. 20231	
NISHI PASANYA	
Name of person mailing paper and fee	Signature of person mailing paper and fee

METHOD AND SYSTEMS FOR SCREENING CHINESE ADDRESS DATA

Inventors: Marie Mung Wah Low
Blk 303 #07-529 Clementi Avenue 4
Singapore 120303, Malaysia

Xin Yu
No. 68, West Hexiang Road
Xiamen, China

Albert Cheh Wooi Tan
8 Medan Nangka, Desa Damai, 14000 Bukit Mertajam
Penang, Malaysia

Zhen Ming Xi
Room 1904 Block 3, Lane 535 West Long Hua Road
Shanghai 200232, China

Loy Lee
Block E-3-6, Taman Desa Relau, Lebuhr Relau 2
11900 Penang, Malaysia

Soon Tat Chan
2-13A-5 Sri Pangkor Condominiums, Jalan Pangkor
Penang 10050, Malaysia

Assignee: Dell Products L.P.
One Dell Way
Round Rock, Texas 78682-2244

David L. McCombs
HAYNES AND BOONE, L.L.P.
901 Main Street
Suite 3100
Dallas, Texas 75202-3789
(214) 651-5533

20090315 0900

Docket No.: DC-02942 (16356.608)
Customer No. 000027683

EXPRESS MAIL NO.: EL 909639427 US DATE OF DEPOSIT: 2/22/02

This paper and fee are being deposited with the U.S. Postal Service Express Mail Post Office to Addressee service under 37 CFR §1.10 on the date indicated above and is addressed to: Box PATENT APPLICATION, Commissioner for Patents, Washington, D.C. 20231

NISHA ARYA [Signature]

Name of person mailing paper and fee Signature of person mailing paper and fee

Background

The present disclosure relates to methods and systems for comparing two databases of Chinese language items. In particular, the disclosure is applicable to permit comparison of items which are data such as addresses of individuals and/or organizations.

Multiple standards exist for writing Chinese text. Besides traditional Chinese character sets (which remain in widespread use in regions such as Taiwan and Hong Kong), texts in the People's Republic of China are written in Simplified Mandarin characters. Furthermore, Chinese may be transcribed into the Roman alphabet as "Pin Yin characters", or by other systems, such as the system defined by the ALA-LC romanization tables.

Conversion between the various standards is common. For example, one conventional order management system (SMARTS) requires that billing and shipping addresses are keyed in using Pin Yin characters, and the Pin Yin characters are then converted into double byte Simplified Mandarin characters for storage in the SMARTS database. Note, however, that not all conversions are unambiguous. For example, a single Simplified Mandarin character can correspond (in Pin Yin) to any of several sets of Roman letters. Similarly, a single set of Roman

letters (in Pin Yin) may correspond to multiple Simplified Mandarin characters, and these Simplified Mandarin characters will have different meanings.

Since Chinese text in different databases may be stored using different standards, comparing the items in different databases is a difficult process. For example, the US Government has issued a "Denied Parties List" ("DPL") and transactions with parties on the list are forbidden. This list is only published in English (i.e. a mixture of conventional English words and transliterations into Roman letters of Chinese words) and there is no indication that in the future it will be translated into Simplified Mandarin Characters. For this reason it is difficult to compare the list with the names stored in an order management system such as SMARTS.

The difficulty of comparing the two lists leads to a risk that a supplier of products will erroneously supply products to parties on the DPL, leading to violations of the US Export Regulations. Such violations carry steep penalties which include, but are not limited to, monetary fines on the exporter (a corporation and/or individuals), possible imprisonment or denial of export privileges.

Summary

The present disclosure attempts to address the above problem, and in particular to provide methods and systems for comparing two databases which each include Chinese text data items such as addresses of entities which are individuals or organizations, and which employ different Chinese writing systems for the Chinese text data items.

In general terms the present disclosure proposes that the Chinese text items of both databases are converted into a common standard language, particularly the

Pin Yin transliteration standard. In the conversion process, any items which may be converted in multiple ways are converted in each of those ways. The items in the two converted databases are then compared.

Specifically, a first aspect of the disclosure is a computer-implemented method for comparing two databases which each comprise Chinese text data items specifying addresses, the method comprising:

for each of the databases, converting any of the Chinese text data items which are not in a predefined common Chinese language format into that common format, any items in at least a first of the databases which are convertible into the common format in multiple ways being converted in all those ways to generate items in the common format; and

comparing the data items in the common format, to identify Chinese text data items in the first database corresponding to Chinese text data items in the second database.

In a second aspect the present disclosure proposes a computer system for comparing two databases which each comprise Chinese text data items specifying addresses, the computer system comprising:

a first conversion unit for converting the Chinese text data items of a first of the databases into a predefined common Chinese language format, any items in a first database which are convertible into the common format in multiple ways being converted in all those ways to generate items in the common format;

a second conversion unit for converting the Chinese text data items of the second of the databases into the common Chinese language format, and

a comparison unit for comparing the converted data items to identify the Chinese text data items in the first database which correspond to Chinese text data items in the second database.

PATENT

Docket No.: DC-02942 (16356.608)

Customer No. 000027683

Note that if the data items in the second database are already in the common format, then the second conversion unit may be omitted.

The common Chinese language format is preferably Pin Yin characters. The first database may be an order management system having data items including shipping and/or billing address. The Chinese text data items in the first database may be in Simplified Mandarin characters. The second database may be in English or a combination of conventional English words and Pin Yin. For example, the second database may be some or all of the Denied Parties Listing issued by the US government.

"Chinese text data items" may be defined as the items which are in a Chinese language, such as Mandarin. Alternatively or additionally, "Chinese text data items" may be defined to include, or consist of, data items associated with addresses within designated Chinese territory, such as the People's Republic of China (which may be defined here to include, or to exclude, the territory of Hong Kong) and/or optionally any other territories where a Chinese language is in common use for billing and/or shipping (particularly one where Simplified Mandarin Characters are in common use).

Note that either database may, in addition to the Chinese text data items, include items which are not Chinese text data items. For example, the order management database may include data relating to parties which have no connection to China. Similarly, the second database (in the case that it is some or all of the DPL) includes items identifying entities for which the address is not within a specified Chinese territory. Preferably in each case, the conversion process only converts the Chinese items in each of the databases, and the comparison

determines whether the converted items of the first database correspond to any of the converted items of the second database.

Brief Description of the Drawings

Further advantages and features of the disclosure will be discussed in relation to an embodiment which is described, for the sake of example only, with reference to the following figures in which:

Fig. 1 is a block diagram illustrating a method which is an embodiment of the present disclosure;

Fig. 2 is a block diagram of the structure of a system which is an embodiment of the present disclosure, and which performs the method of Fig. 1;

Fig. 3 shows a window presented by the system of Fig. 2, and used to generate addresses in Simplified Mandarin characters from Pin Yin characters;

Fig. 4 is a window presented by the system of Fig. 2 and showing the addresses stored in the system in Simplified Mandarin characters;

Fig. 5, which is composed of Figs. 5(a) to 5(c), shows the steps of converting Simplified Mandarin Characters to Pin Yin characters in the method of Fig. 1;

Fig. 6 shows the database of Pin Yin characters generated from the DPL by the method of Fig. 1; and

Fig. 7 is a window presented by the system of Fig. 2 showing the result of a comparison of two databases.

Detailed Description

Figure 1 shows the steps of a method according to an embodiment for comparing the addresses of potential recipients of goods with at least part of the Denied Parties Listing (DPL). The method is performed by the system shown in Fig. 2.

The system of Fig. 2 comprises an order management system 100, such as the SMARTS system, including a database 110 for storing shipping and/or billing addresses of individuals and/or companies which have placed orders or which are due to receive orders, and a data input device 120 for entering data using Pin Yin characters into the database 110. Only one data input device 120 is shown, but in practice there may be multiple such units.

The system further includes a second database 130 for storing the English-language DPL.

The system further includes a first conversion unit 140 for converting the Simplified Mandarin data items in the first database 110 into Pin Yin data items to form a first Pin Yin database 150. This process does not erase the database 120.

The system further includes a second conversion unit 160 for converting the English language data in the second database 130 into Pin Yin data items in a second Pin Yin database 170. This process does not erase the second database 130.

Finally, the system includes a comparison unit 180 for comparing the Pin Yin items in the first and second databases 150, 170, and an output unit 190 for notifying an operator of the system of any matches between items in the first and second Pin Yin databases 150, 170 which are discovered by the comparison unit 180.

The first two steps of the method of Fig. 1 (i.e. the ones above the dashed line in Fig. 1) are the known steps of entering data into the first database 110 of the order management system 100. Specifically, in step 10 users such as inside sales representatives use the data input devices 120 to enter data such as billing and shipping addresses into the order management system 100.

A window presented to the user by the order management system 100 is shown in Fig. 3. Using this window, in step 20, and helped by user intervention, the order management system 100 converts the input data into Simplified Mandarin double byte characters, to form items in the first database 110. When items from the first database 110 are printed out they are in Simplified Mandarin, as is generally required for use on shipping and invoice documents. Fig. 4 shows an element from the second database, having the whole of the billing and mailing addresses written in double byte Simplified Mandarin characters. Note that the database 110 may contain further items which are not Chinese-related, and which are not relevant to the present disclosure. Such items, if they are already in the English language, may be compared directly with items (e.g. non-Chinese items) in the database 130 by known methods.

In step 30, the billing and shipping data which resides in the first database 110 in Simplified Mandarin double byte form is converted by the first conversion unit 140 into Pin Yin characters, to form items in the first Pin Yin database 150. As noted above, a single Simplified Mandarin character may correspond to multiple sets of Pin

Yin characters, and these sets of Pin Yin characters will have different meanings. Hence, the first conversion unit 140 generates, for each Simplified Mandarin item in the first database 110, ALL the possible sets of Pin Yin characters which can be derived from that item, and each of these sets of Pin Yin characters forms an item in the database 150. We have determined that this "simplistic" process does not, however, compromise the integrity of the screening process.

Specifically, the conversion carried out in step 30 by the conversion unit 140 may be performed using a conversion file such as the default copy of the loaded Microsoft Windows 98 Simplified Chinese Operating System. The default file system location for each install can be found at c:\windows\system\winpy.com of each PC into which this operating system is installed.

Fig. 5 shows an example of the process of step 30. The address displayed in the window of Fig. 4 is order no. 4602249011 in the first database, as shown in Fig. 5(a). Fig. 5(b) shows the various ways in which each of the Simplified Mandarin characters can be converted into Pin Yin. Most only have one Pin Yin version, but three of them have two Pin Yin transliterations, of which one is shown shaded.

Using the table of Fig 5(b), the string of Simplified Mandarin characters in converted into a string of Pin Yin characters. Each Simplified Mandarin character with multiple Pin Yin representations is converted as one representation followed by the other representation(s). This string is shown in Fig. 5(c) by indicating a first Pin Yin representation for each such Mandarin character followed by the other Pin Yin representation shaded.

In step 40, the Chinese addresses in the second database 130 are converted into Pin Yin by the second conversion unit 160 to form the items of the second Pin

Yin database 160. Note that this conversion process must normally be performed manually by a Chinese speaking operator, though the process may in principle also be automated or semi-automated.

Fig. 6 illustrates the conversion operation. Each row corresponds to an entity on the DPL (labelled PIN_YIN_1 up to PIN_YIN_9). For example, the entity PIN_YIN_2 is the "Beijing Institute of Structure and Environmental Engineering". The US Government DPL includes an address for this entity of "No. 36 Wanyuan Road Beijin China (PRC)" (this address is labelled "BXA DPL address" in Fig. 6). Note that the address is a mixture of conventional English words (e.g. "Road") and Pin Yin (e.g. "Wanyuan"). In step 40, the BXA DPL address is converted (e.g. by an operator) into a wholly Pin Yin address. For reference, the corresponding Simplified Chinese Address is shown in the right hand column of Fig. 6, though the generation of this column is not necessary to the present disclosure.

While in principle it would be possible to convert all the items in the DPL into Pin Yin, the present embodiment only converts the addresses of the Chinese items in the DPL. For example, "Chinese" in this context may be defined as the items which are addresses in the People's Republic of China and optionally other territories. By taking this "simplistic" approach, the number of conversions (and thus of subsequent comparisons) is much reduced. In general, this does not reduce the integrity of the screening, since the screening process is based on addresses, and addresses by their nature are not "mobile".

In step 50, a comparison is performed of the first and second Pin Yin databases 150, 170 to determine matches. This done by automatically extracting matches between the Pin Yin strings in the first database (e.g. the string shown in

Fig. 5(c)), and the Pin Yin strings in the second database (the "Pin Yin addresses" column of Fig. 6.

Fig. 7 shows a window optionally presented to the user by comparison unit 180 for the user to decide how the match is to be treated. As shown, a possible match has been found between order number 402211081 (shown in Figs. 4 and 5, and in the upper part of Fig. 7) and entity PIN-YIN_4 in the list of Fig. 6 (shown in the lower part of Fig. 7). Note that the entity name in the DPL ("Beijing Aerospace Automatic Control Limited") is different from the name ("DaLi Furniture (China) Ltd.") in which the order was made; the embodiment has found the match based on the addresses alone. By entering ticks in appropriate option boxes in the window of Fig. 7 and then clicking on "OK", the user can indicate how the match is to be treated.

Step 50 may if desired be performed by a DPL compliance department of the organization operating the order management system. The matches can be incorporated into a local DPL , i.e. a list of parties (not necessarily the same as those on the US government's DPL) with which the organization operating the order management system refuses to transact business, at least without a screening operation. The local DPL may be subsequently used to add to an export management system for export compliance screening purposes as well as for the generation of export/shipping documents.

Thus, steps 30 and 40 have resulted in a common platform (Pin Yin), enabling in step 50 the compliance screening of addresses of China orders.

The embodiment may be operated in a batch mode in which a plurality of items in the first database 110 (e.g. all the Chinese items in the first database 110) are converted into Pin Yin items one after another (e.g. as a continuous sequence)

to form the database 150, and later each of the converted items in the database 150 are compared (e.g. one after another) with the converted items of the second database 170.

Alternatively, step 30 may be performed for the items of the first database 110 individually (for example, whenever a new item is added to the first database 110), and step 50 may be performed for the resultant items in the database 150 by comparing the individual converted items with all the converted items of the second Pin Yin database 170. If no matches are found, the contents of the database 150 may be discarded. In other words, in this variant of the embodiment, the first Pin Yin database 150 need not contain at any time more than the number of Pin Yin items which are derived from a single one of the Simplified Mandarin items in the database 110.

The comparison in step 50 may be performed as described above. If any matches are found, the output unit 190 is used to notify an operator of the system, who may cancel the corresponding order. Alternatively, though less preferably, the order may be cancelled automatically.

Although illustrative embodiments have been shown and described, a wide range of modification, change and substitution is contemplated in the foregoing disclosure and in some instances, some features of the embodiments may be employed without a corresponding use of other features. Accordingly, it is appropriate that the appended claims be construed broadly and in a manner consistent with the scope of the embodiments disclosed herein.